My Tennis Story

As said by Albert Einstein, "*Not everything that can be counted counts".* So it is significant to find out which data is important and which is not. In the case study here, revolving around the last 10 year's finals of the Australian Open, we are studying the deciding factors in the tennis game. Tennis, or for that matter any other game, has a lot of factors that tell about competencies of the winner over the loser. This is an attempt to figure out the importance of few of the most frequently used statistics in tennis, and how they can be used to predict the winner as the match proceeds.

For this, we are taking into consideration 4 of the most widely used parameters of the tennis world. Service-points-win% & return-point-win% are considered as they contribute the most to the overall score, and 70+% of the match is composed of traditional rallies than faults and breaks. They sum up to form the total score, which eventually decides the winner. We have also taken double faults in consideration because they are an integral part of the game and serve as the game changer by affecting mental states of the players. Breakpoints on the other hand help in breaking the serve and winning the game. Also, the server at the breakpoint has an advantage, especially if it is a double or triple breakpoint. If both players are of same competency (eg. both are good servers), breakpoints usually decide the victory as the server has upper hand.

Going ahead with the data that is being referred to, it can be seen that both the service-point-win% and return-point-win% stats are clearly favoring the winner. In 90% of the cases, it can be seen that the winner has been consistently scoring more return and service points. This also makes perfect sense in line with the actual gameplay. Since faults, breakpoints, etc. are uncommon scenarios in the game, these points are what actually affect the bulk of the score. Similarly, if we see the number of double faults made by both the players, the loser always has around 1.5 times more double faults. This emphasizes the fact mentioned above that faults can create a confidence boost up in the rival, while decreasing the faulter's self-confidence. He will then start playing defensively, giving full chance to the opposition to play attack. If we see the 2011 statistics, the loser performed very well on the breakpoints but still could not match up to the opponent who flawlessly played without any faults. Moving on to breakpoint conversion%, it can be seen that most of the times the winner would have scored more conversions. Although, it is not a clear cut metric because in some matches loser has not been able to win despite successful conversions. This can be explained with the fact that breakpoints are not a general scenario. The situation occurs very less and even if a player is able to get the break point, it is just a single point after all. It will not be able to affect the overall score to a great extent if the player's performance is not consistent.

The visualization that I have used is to depict the essence of Australian Open tournament so that the viewer feels as though he is seeing the progress of the matches from within the stadium. A background picture of the hard court and backgrounds of the overlaid graphs in similar color supports my idea. Out of the three graphs depicted, I wanted to focus my discussion and observations more on the points won by the players, hence the intention of placing this graph at the top. Color coding for the data used shows a clear distinction between the winners and losers: winners in the hues of green and blue which are positive colors, losers in the shades of red to show negative symbolism. The lines and dots clearly show the winners' performance over the opponents over the years, easily showing the winners have the upper hand (here the upper line). Similarly, for breakpoint conversion% graph, the colors are used in a similar fashion. As to the bar chart, the idea was to show how far ahead (here to the right) do the players go and succeed in converting the service breaks. Lastly, we have the double faults graph. Here, even for the winners, I have used the color in the lines of red although a lighter shade, pertaining to the data - double fault - which is after all not a positive act. Losers are assigned a darker shade in comparison. It is easier to see the ratio of faults made by each of the parties, and if you want numbers, they can be seen right on top of the bars.

All the above graphs have been created with the help of Google Sheets. And to sum up from the above, I would say that the major factor in deciding the winner is the one who has consistently scored more service/return points. However, when players are of equal level, breakpoints and faults can be a game changer.